Binarizing Sparse Convolutional Network for Efficient Point Cloud Analysis

Xiuwei Xu[1,2], Ziwei Wang[1,2], Jie Zhou[1,2], Jiwen Lu[1,2]

[1] Department of Automation, Tsinghua University, China

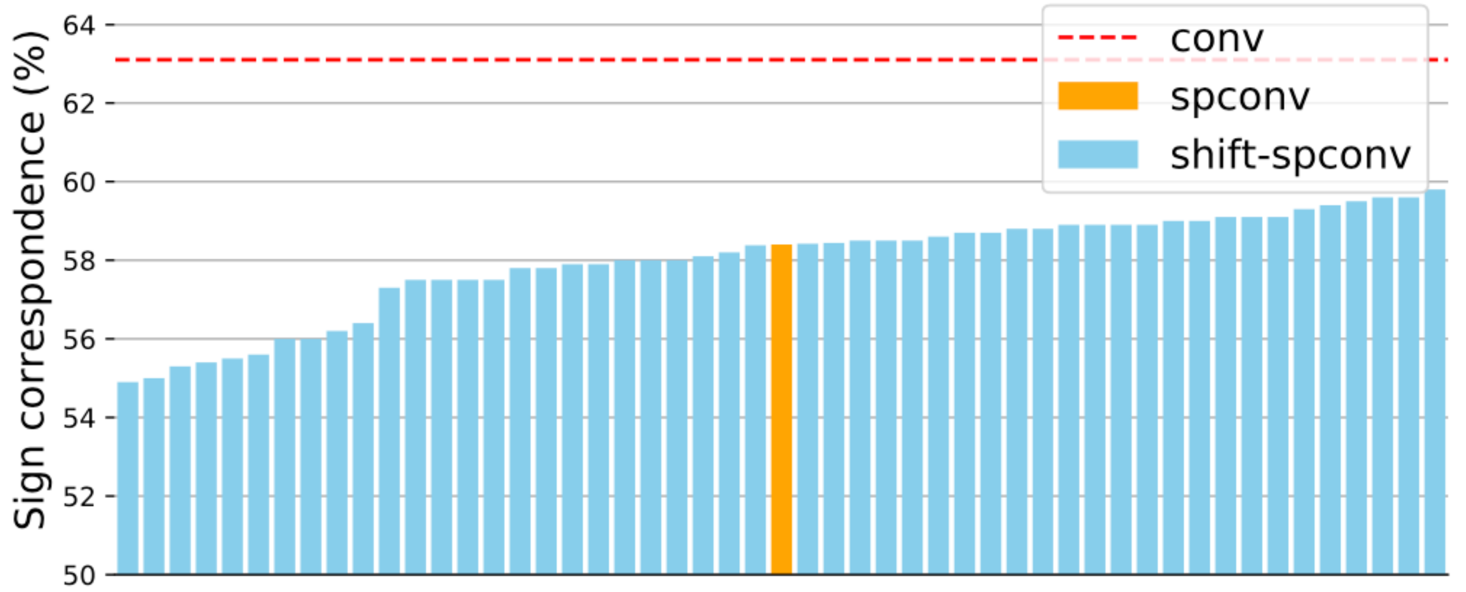
[2] Beijing National Research Center for Information Science and Technology, China

Tsinghua University

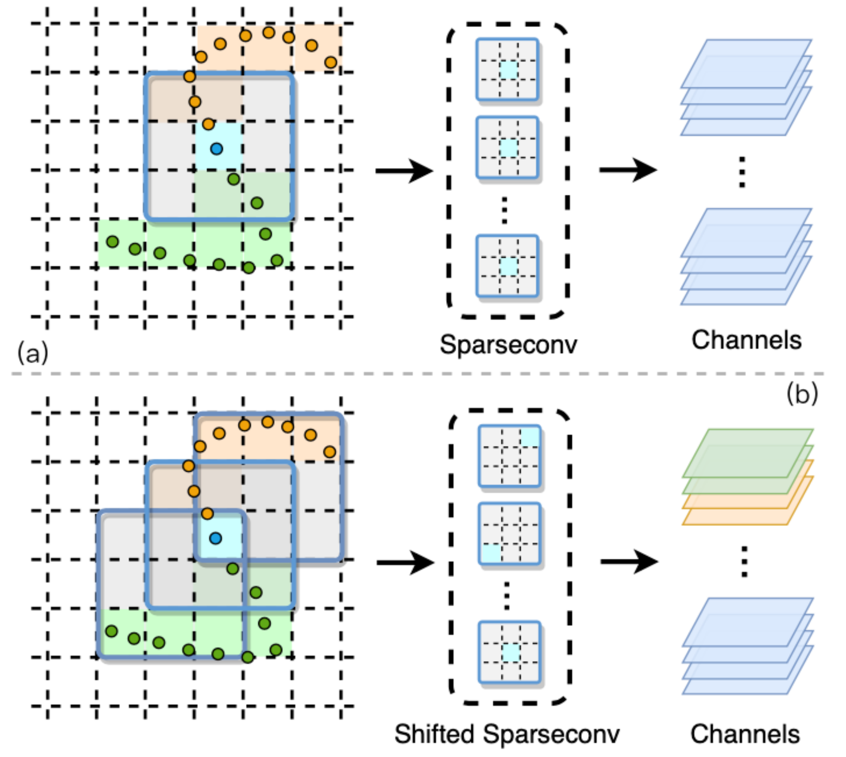
JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Introduction

- **Problem Definition:** As a basic operator in 3D point cloud analysis, sparse convolution is widely adopted in most state-of-the-art architectures for point cloud analysis and so it is desirable to further improve its efficiency for more practical application.
- We empirically find **sparse convolution operation brings larger quantization errors compared to standard convolution**. Here we calculate the sign correspondence of activations for the first binary layer in convolutional and sparse convolutional networks.



- **Motivation:** For a single active site, a 3 × 3 × 3 convolution kernel will operate 27 times while sparse convolution kernel only operates at the center. What if we keep the same number of operations with sparse convolution but operates at other location? To answer it, we extend sparse convolution to enable it to active at different locations and propose shifted sparse convolution operator (SFSC).
- We find **different SFSC layers vary a lot in quantization errors and a proportion of them are more robust to binarization** compared to sparse convolutional layer.
- In another word, if we can find out the optimal configurations for all SFSC layers in a network, the quantization error can be reduced without additional computational cost.



Approach

- **Preliminary:** sparse convolution can be formulated as:

$$F_0(\boldsymbol{W}, \mathbf{x_u}) = \sum_{\mathbf{i} \in N^D(\mathbf{u})} \boldsymbol{W}_\mathbf{i} \mathbf{x}_{\mathbf{u}+\mathbf{i}}$$

which skips the non-active regions that only operates when the center of convolutional kernel covers active voxels.

- **Shifted Sparse Convolution:** we extend sparse convolution to allow it operates at locations other than the center of kernel:

$$F_k(\boldsymbol{W}, \mathbf{x_u}) = \sum_{\mathbf{i} \in N^D(\mathbf{u}+s_k)} \boldsymbol{W}_\mathbf{i} \mathbf{x}_{\mathbf{u}+\mathbf{i}}$$

$$s_k \in \mathbb{R}^3,\ k \in \{1, 2, ..., n_s\}$$

- In a SFSC layer, instead of applying the same sparse convolution operation for all output channels as in a general sparse convolutional layer, we uniformly divide the output channels into several groups (namely channel group), each with a specific SFSC operation. It can be formulated as:

$$y = \mathrm{concat}(f_1(\boldsymbol{W}_1, x), ..., f_{n_g}(\boldsymbol{W}_{n_g}, x)),\ f_i \in \boldsymbol{F}_{n_s}$$

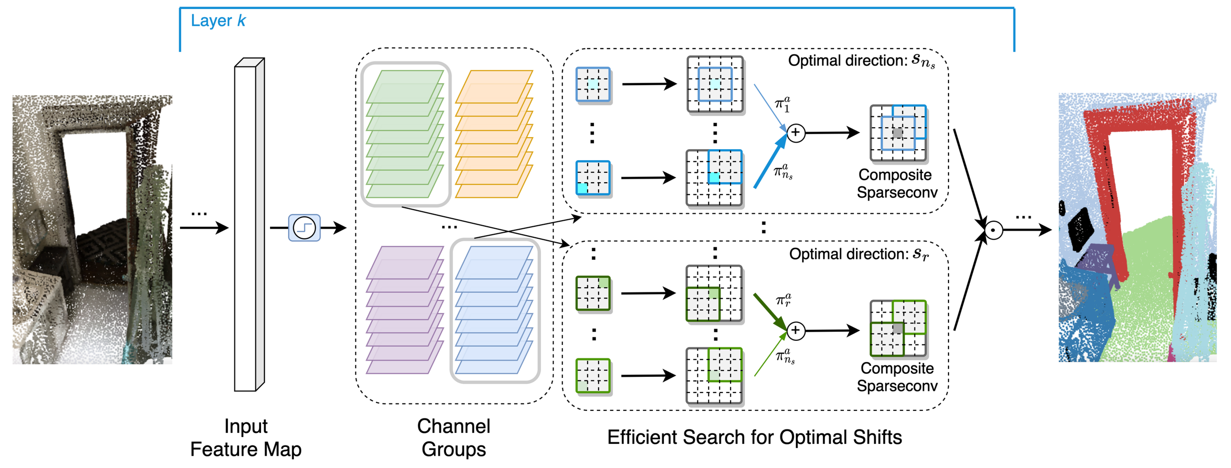
- **Efficient Search for Shift Operation:** To find out the (near) optimal configurations for all SFSC layers in a network, we formulate this by searching the optimal SFSC operation in each channel group:

$$f_i = \sum_{j=1}^{n_s} o_{ij}^a F_j,\ i \in \{1, 2, ..., n_g\} \quad \text{s.t.} \sum_j o_{ij}^a = 1,\ o^a \in \{0, 1\}.$$

$$\Rightarrow$$

$$f_i^* = \sum_{j=1}^{n_s} \pi_{ij}^a F_j,\ i \in \{1, 2, ..., n_g\} \quad \text{s.t.}\ \pi^a \in [0, 1],\ \pi_{ij}^a = \frac{1}{1 + \exp(-\alpha_{ij})}$$

- **Overall framework:**



Experiments

- **Main results:**

	Method	Param.	5cm voxel OPs	5cm voxel mIoU	5cm voxel mAcc	2cm voxel OPs	2cm voxel mIoU	2cm voxel mAcc
UNET-S	Real valued	4.335	1.21×10^9	65.2	73.3	5.32×10^9	68.7	78.5
	XNOR-Net	0.136	8.07×10^7	33.3	38.9	3.79×10^8	21.0	26.1
	XNOR-Net++	0.136	8.07×10^7	12.6	15.9	3.79×10^8	11.2	13.7
	BiPointNet	0.136	8.07×10^7	30.1	36.2	3.79×10^8	18.4	20.7
	Bi-Real-Net	0.138	8.12×10^7	48.3	56.6	3.82×10^8	51.2	63.3
	ReActNet	0.138	8.12×10^7	43.6	50.2	3.82×10^8	46.9	52.9
	BSC-Baseline	0.139	8.12×10^7	51.7	61.8	3.82×10^8	54.9	65.3
	BSC-Manual	0.139	8.12×10^7	53.2	63.7	3.82×10^8	57.8	66.6
	BSC-Net	0.139	8.12×10^7	54.4	65.2	3.82×10^8	61.4	70.4
UNET-H	Real valued	30.104	7.65×10^9	67.6	75.1	3.38×10^{10}	71.0	79.0
	XNOR-Net	0.939	3.61×10^8	46.6	53.5	1.75×10^9	34.9	40.1
	XNOR-Net++	0.939	3.61×10^8	13.3	16.4	1.75×10^9	12.8	16.2
	BiPointNet	0.939	3.61×10^8	45.2	52.4	1.75×10^9	34.3	39.8
	Bi-Real-Net	0.948	3.63×10^8	53.4	63.2	1.76×10^9	57.3	66.9
	ReActNet	0.949	3.63×10^8	52.2	59.0	1.76×10^9	57.1	67.0
	BSC-Baseline	0.952	3.63×10^8	56.0	65.9	1.76×10^9	59.3	68.3
	BSC-Manual	0.952	3.63×10^8	59.3	68.7	1.76×10^9	62.2	70.1
	BSC-Net	0.952	3.63×10^8	62.2	70.5	1.76×10^9	63.9	71.6

	Method	Param.	OPs	mIoU	mAcc	Acc
FCN-S	Real valued	2.496	1.24×10^9	33.9	47.7	64.7
	XNOR-Net	0.108	1.72×10^8	22.1	32.7	57.3
	XNOR-Net++	0.108	1.72×10^8	8.5	13.5	43.9
	BiPointNet	0.108	1.72×10^8	24.9	35.7	59.3
	Bi-Real-Net	0.110	1.75×10^8	27.3	38.4	60.0
	ReActNet	0.110	1.75×10^8	25.4	36.6	58.9
	BSC-Baseline	0.110	1.75×10^8	27.8	39.9	60.1
	BSC-Manual	0.110	1.75×10^8	28.7	40.9	60.2
	BSC-Net	0.110	1.75×10^8	29.7	42.1	61.2
FCN-H	Real valued	10.025	4.82×10^9	36.9	50.4	67.2
	XNOR-Net	0.357	3.08×10^8	27.1	38.8	59.6
	XNOR-Net++	0.357	3.08×10^8	8.4	13.6	43.2
	BiPointNet	0.357	3.08×10^8	28.1	40.8	60.1
	Bi-Real-Net	0.361	3.09×10^8	30.4	41.8	61.1
	ReActNet	0.361	3.09×10^8	27.0	40.1	58.9
	BSC-Baseline	0.362	3.09×10^8	32.0	44.4	63.3
	BSC-Manual	0.362	3.09×10^8	32.5	45.1	63.5
	BSC-Net	0.362	3.09×10^8	33.9	46.2	64.5

- **Ablation study:**

Search space	Group number	mIoU(%)	mAcc(%)
Baseline: $\{S\}$	–	51.7	61.8
$\{S, C_1, C_8\}$	8	53.6	63.9
$\{S, C_1, C_4, C_6, C_7\}$	8	54.2	64.9
$\{S, C_1, C_2, C_3, C_4, C_5, C_6, C_7, C_8\}$	2	52.9	63.8
	4	53.4	64.6
	8	**54.4**	**65.2**
	16	54.0	64.5

Relaxation	Derivation	mIoU(%)	mAcc(%)
Random	$D(32,32) \rightarrow D(1,1)$	51.5	61.2
Softmax	$S(32,32) \rightarrow S(1,1) \rightarrow D(1,1)$	51.8	62.0
	$S^*(32,32) \rightarrow S(1,1) \rightarrow D(1,1)$	52.3	63.2
	$S(1,1) \rightarrow D(32,32) \rightarrow D(1,1)$	53.5	64.5
Sigmoid	$S(32,32) \rightarrow S(1,1) \rightarrow D(1,1)$	52.6	64.0
	$S^*(32,32) \rightarrow S(1,1) \rightarrow D(1,1)$	53.7	64.7
	$S(1,1) \rightarrow D(32,32) \rightarrow D(1,1)$	**54.4**	**65.2**

- **Visualization of different methods:**